DATA SCIENCE INTERVIEW QUESTIONS

1 WHAT IS DATA SCIENCE?

Data Science deals with the processes of data mining, cleansing. analysis, visualization, and actionable insight generation. Data Science is the mining and analysis of relevant information from data to solve analytically complicated problems. It is most widely used technique between Artificial Intelligence and Machine Learning. For example, when you logged on any e-commerce website and browsed some categories and products before purchase, you are generating data, which will be helpful for analysts to know your behavior about purchase.

2 WHAT ARE RECOMMENDER SYSTEMS?

A subclass of data sifting frameworks that are intended to anticipate the inclinations or evaluations that a client would provide for an item. Recommender systems are generally utilized in music, pictures, research, news, articles, social labels, and so on.

3 CAN YOU UTILIZE MACHINE LEARNING FOR TIME SERIES ANALYSIS?

Yes machine learning can be utilized for time series analysis but it depends on the applications.

4 HOW WILL YOU ASSESS THE STATISTICAL IMPORTANCE OF AN INSIGHT?

By utilizing Hypothesis Testing, we can assess the statistical significance of an insight.

5 FOR TEXT ANALYTICS, PYTHON OR R WHICH ONE WOULD YOU GIVE THE PREFERENCE?

Python is the best choice for text analytics as it has Pandas library that provides easy to use data structures and better performance data analysis gadgets.

6 WHICH METHOD IS UTILIZED TO FORECAST CATEGORICAL RESPONSES?

Supervised machine learning i.e. Classification technique is widely utilized in mining for classifying data sets.

7 WHAT ARE THE BASIC EXPECTATIONS TO BE MADE FOR LINEAR REGRESSION?

Statistical independence of errors, normality of error distribution, linearity and additivity.

8 WHAT IS THE DIFFERENCE BETWEEN DATA SCIENCE AND MACHINELEARNING?

Data Science deals with the processes of data mining, cleansing, analysis, visualization, and actionable insight generation, whereas, machine Learning is the part of Data Science which enables the system to process datasets autonomously without any human interference by utilizing various algorithms to work on massive volume of data generated and extracted from numerous sources.

9 WHAT IS THE FORMULA TO CALCULATE R-SQUARE?

R-Square can be calculated as:

1- (Residual Sum of Squares/ Total Sum of Squares)

10 WHAT BASIC KNOWLEDGE REQUIRED FOR DATA SCIENTIST?

Data Scientist must have the basic knowledge of mathematics, computer programming and statistics to solve the complex data problems in an efficient way to boost the business revenue.

11 NAMES OF BASIC MODELS OF MACHINE LEARNING?

There are two basic models of Machine learning are:
Supervised Machine Learning
Unsupervised Machine Learning

12 DO YOU KNOW ABOUT INTERPOLATION AND EXTRAPOLATION?

Interpolation is assessing a value from two known values from a list of values, whereas, extrapolation is assessing a value by extending a known set of values or evidences.

13 WHAT ARE THE BASIC BENEFITS OF DATA SCIENCE?

Data Science helps in finding and refining of target viewers. It ensure better communication between service providers and service utilizers. Also improved business value and better risk analysis

14 DO YOU KNOW POWER ANALYSIS?

Power Analysis is an experimental design method for determining the effect of a given sample size.

15 WHAT IS COLLABORATIVE FILTERING?

It is used by the recommender systems to find patterns or information by collaborating viewpoints, several data sources and various agents.

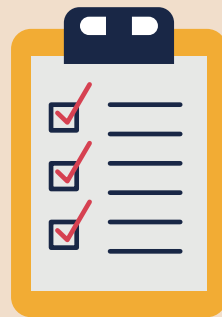**16 WHAT ARE THE TOP TOOLS UTILIZED IN DATA SCIENCE?**

R (a language for statistical computing and graphics)
Python
Tableau
Keras
Jupyter Notebook

17 WHAT ARE THE MAIN PROCESS OF DATA SCIENCE?

Data Exploration:
Modeling: Model Testing:
Model deployment: